



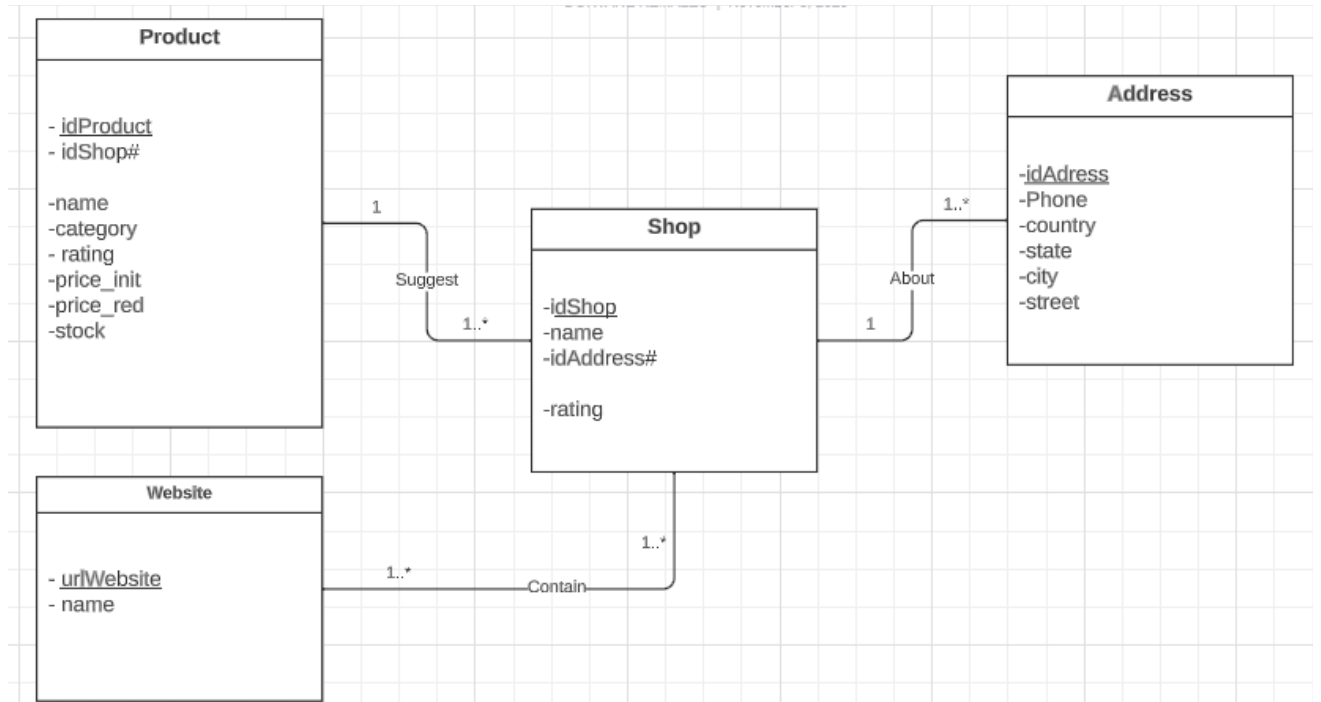
AFRICA DIGITAL XPERTS  
S O L U T I O N S

PROJECT

**E-COMMERCE DATA SCRAPING & ANALYTICS**

# MODELISATION DES DONNEES

## Modélisation conceptuelle



Modélisation logique en considérant un SGBD SQL : MySql

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Views

Stored Procedures

Functions

sys

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

Indexes in Table

Visible	Key	Type	Uniq	Columns
<input checked="" type="checkbox"/>	PRIMARY	BTREE	YES	idWebsite
<input checked="" type="checkbox"/>	url	BTREE	YES	url

Index Details

Key Name: PRIMARY

Index Type: BTREE (clustered)

Allows NULL:

Cardinality: 3

Comment:

User Comment:

Packed:

Unique: YES

Columns in table

Column	Type	Nullat	Indexes
idWebsite	int	NO	PRIMARY

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Views

Stored Procedures

Functions

sys

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

Indexes in Table

Visible	Key	Type	Uniq	Columns
<input checked="" type="checkbox"/>	PRIMARY	BTREE	YES	idShop
<input checked="" type="checkbox"/>	idAddress	BTREE	NO	idAddress

Index Details

Key Name: PRIMARY

Index Type: BTREE (clustered)

Allows NULL:

Cardinality: 0

Comment:

User Comment:

Packed:

Unique: YES

Columns in table

Column	Type	Nullat	Indexes
idShop	int	NO	PRIMARY
idAddress	int	YES	idAddress
rating	float	YES	

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

SCHEMAS

Filter objects

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Views

Stored Procedures

Functions

sys

Query 1 Address Scraped\_ADXS.Address Scraped\_ADXS.Contain Scraped\_ADXS.Product

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

Indexes in Table

Visible Key	Type	Uniq	Columns
PRIMARY	BTREE	YES	idProduct
idShop	BTREE	NO	idShop

Index Details

Key Name: PRIMARY

Index Type: BTREE (clustered)

Allows NULL:

Cardinality: 0

Comment:

User Comment:

Packed: Unique: YES

Columns in table

Column	Type	Nullat	Indexes
idProduct	int	NO	PRIMARY
idShop	int	YES	idShop
name	varchar(200)	YES	
category	varchar(100)	YES	
rating	float	YES	
pricelnit	float	YES	
priceRed	float	YES	

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

SCHEMAS

Filter objects

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Views

Stored Procedures

Functions

sys

Query 1 Address Scraped\_ADXS.Address Scraped\_ADXS.Contain

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

Indexes in Table

Visible Key	Type	Uniq	Columns
PRIMARY	BTREE	YES	idContain
idShop	BTREE	NO	idShop
idWebsite	BTREE	NO	idWebsite

Index Details

Key Name: PRIMARY

Index Type: BTREE (clustered)

Allows NULL:

Cardinality: 0

Comment:

User Comment:

Columns in table

Column	Type	Nullat	Indexes
idContain	int	NO	PRIMARY
idShop	int	YES	idShop
idWebsite	int	YES	idWebsite

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

Query 1 Address Scraped\_ADXS.Address Scraped\_ADXS.Contain

SCHEMAS

Filter objects

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Views

Stored Procedures

Functions

sys

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

Indexes in Table

Visible	Key	Type	Uniq	Columns
✓	PRIMARY	BTREE	YES	idAddress

Index Details

Key Name: PRIMARY

Index Type: BTREE (clustered)

Allows NULL:

Cardinality: 0

Comment:

User Comment:

Packed: Unique: YES

Columns in table

Column	Type	Nullat	Indexes
idAddress	int	NO	PRIMARY
phone	varchar(45)	YES	
country	varchar(45)	YES	
state	varchar(45)	YES	
city	varchar(45)	YES	
street	varchar(45)	YES	

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

Query 1 Address Scraped\_ADXS.Address Scraped\_ADXS.Contain

SCHEMAS

Filter objects

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Views

Stored Procedures

Functions

sys

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

Indexes in Table

Visible	Key	Type	Uniq	Columns
✓	PRIMARY	BTREE	YES	idAddress

Index Details

Key Name: PRIMARY

Index Type: BTREE (clustered)

Allows NULL:

Cardinality: 0

Comment:

User Comment:

Packed: Unique: YES

Columns in table

Column	Type	Nullat	Indexes
idAddress	int	NO	PRIMARY
phone	varchar(45)	YES	
country	varchar(45)	YES	
state	varchar(45)	YES	
city	varchar(45)	YES	
street	varchar(45)	YES	

MySQL Workbench

Local instance dkemal

File Edit View Query Database Server Tools Scripting Help

Schemas

Query 1 Address Scraped\_ADXS.Address

SCHEMAS

Filter objects

Scraped\_ADXS

Tables

Address

Contain

Product

Shop

Website

Info Columns Indexes Triggers Foreign keys Partitions Grants DDL

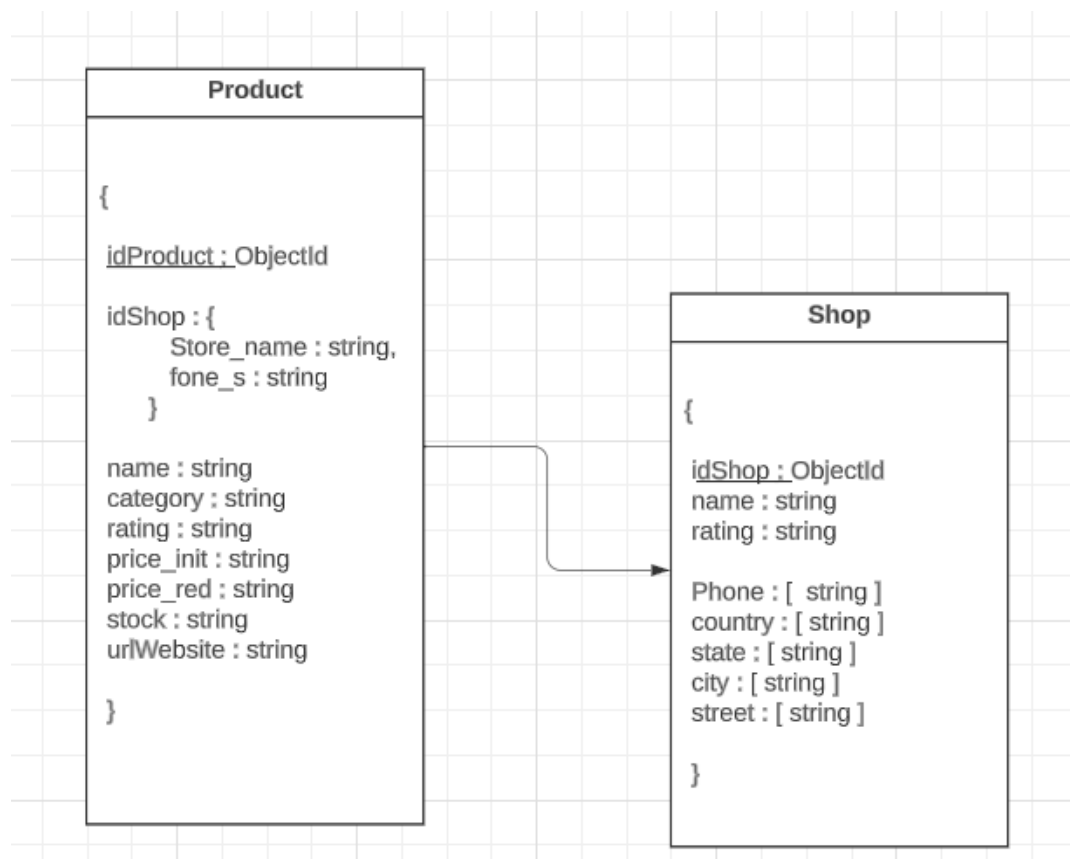
Column	Type	Default Value	Nullat	Character	Collation	Privileges	Extra
city	varchar(45)		YES	utf8mb4	utf8mb4_0900	select,insert,update,referenc	
country	varchar(45)		YES	utf8mb4	utf8mb4_0900	select,insert,update,referenc	
idAddress	int		NO			select,insert,update,referenc	auto_increme
phone	varchar(45)		YES	utf8mb4	utf8mb4_0900	select,insert,update,referenc	
state	varchar(45)		YES	utf8mb4	utf8mb4_0900	select,insert,update,referenc	
street	varchar(45)		YES	utf8mb4	utf8mb4_0900	select,insert,update,referenc	

## Avantages et limites de la bd SQL

Cette base de données est assez simple à concevoir et monter, ACID (Atomicité, Cohérence, Isolation, Durabilité). Mais elle est très rigide, peu évolutive et inadaptée pour les Big Data. Une simple opération de mise à jour/ chargement pourrait être très coûteuse.

## Modélisation logique en considérant un SGBD NoSQL Document : MongoDB

### Visualisation des Collections



# Base de données

MongoDB Compass - localhost:27017/Scraped\_ADXS.Shop

Connect Edit View Collection Help

localhost:27017 Documents Scraped\_ADXS.S...

My Queries

Databases

Scraped\_ADXS

Product

Shop

admin

config

local

Scraped\_ADXS.Shop

36 1 DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Filter Type a query: { field: 'value' } or Generate query Explain Reset Find Options

ADD DATA EXPORT DATA

1 - 20 of 36

```
{
  "_id": ObjectId('65452087a1a5fe698097adfb'),
  "name": "Series TV",
  "site": "durell_market",
  "rating": 0,
  "street": Array(1)
    0: " Cité U"
  "city": Array(1)
    0: " Yaoundé"
  "state": Array(1)
    0: " Cameroun "
  "country": Array(1)
    0: ""
  "phone": Array(1)
    0: "237694569554"
}
```

MongoDB Compass - localhost:27017/Scraped\_ADXS.Product

Connect Edit View Collection Help

localhost:27017 Documents Scraped\_ADXS.Pr...

My Queries

Databases

Scraped\_ADXS

Product

Shop

admin

config

local

Scraped\_ADXS.Product

3.2k 1 DOCUMENTS INDEXES

Documents Aggregations Schema Indexes Validation

Filter Type a query: { field: 'value' } or Generate query Explain Reset Find Options

ADD DATA EXPORT DATA

1 - 20 of 3150

```
{
  "_id": ObjectId('6544d803a1a5fe698097a1aa'),
  "name": "Bijoux",
  "category": "Bijoux",
  "rating": 0,
  "price_init": 3500,
  "price_red": 2500,
  "url": "https://durrellmarket.com/produit/bijoux/"
}
```

```
{
  "shop": Object
    store_name: "Dianacute"
    fone_s: "681310931"
}
```

Object

```
{
  "_id": ObjectId('6544d803a1a5fe698097a1ab'),
  "name": "Berger Allemand",
  "category": "Non classé",
  "rating": 0,
  "price_init": 0
}
```

## Avantages et limites de cette Base de donnée NoSQL

La logique de ce modèle est très proche de nos données scrappées. La bd est plus flexible, et performante pour des requêtes simples en temps réel. Elle supporte une scalabilité horizontale contrairement à la précédente, et est idéale pour nos Big Data.

## DICTIONNAIRE DE DONNÉES

Collection / Table	NomChamp	Type Champ	Description
Poduct	id	ObjectId	Identifiant du document "Produit", généré et traité par le SGBD
	name	String	Nom du produit
	category	String	Catégorie à laquelle appartient le produit
	rating	Float	Moyenne des notation des clients qui apparaît sur le site. elle est notée sur 5
	price_init	Float	Prix du produit (barré), sans la réduction qui apparaît sur le site
	price_red	Float	Prix final du produit (non barré) qui apparaît sur le site
	url	string	Url qui identifie le site sur lequel le produit à été récupéré
	shop	Object	Objet qui contient les premières informations apparentes (store_name,



			fone_s) de la boutique qui propose ce produit
	<b>store_name</b>	string	Nom de la boutique qui propose le produit
	<b>fone_s</b>	string	Numéro de téléphone de la boutique qui propose le produit
<b>Shop</b>	<b>id</b>	ObjectId	Identifiant du document "Produit", généré et traité par le SGBD
	<b>name</b>	string	Nom de la boutique
	<b>site</b>	string	Nom du site sur lequel elle a été récupérée
	<b>rating</b>	string	Moyenne de notation de la boutique sur le site ou elle a été récupérée
	<b>city</b>	[ string ]	Liste des villes où la boutique est présente
	<b>street</b>	[ string ]	Liste des quartiers
	<b>state</b>	[ string ]	Liste des états
	<b>country</b>	[ string ]	Liste des pays
	<b>phone</b>	[ string ]	Liste des contacts à travers lesquels on peut joindre la boutique

**Remarque :** Grâce à la flexibilité des BD NoSQL , les types des champs ne sont pas des contraintes absolues pour former des documents dans nos collections